CrossMark

Maximum Likelihood Threshold and Generic Completion Rank of Graphs

Grigoriy Blekherman[1] · Rainer Sinn[2]

Abstract The minimum number of observations such that the maximum likelihood estimator in a Gaussian graphical model exists with probability one is called the maximum likelihood threshold of the underlying graph G. The natural algebraic relaxation is the generic completion rank introduced by Uhler. We show that the maximum likelihood threshold and the generic completion rank behave in the same way under clique sums, which gives us large families of graphs on which these invariants coincide. On the other hand, we determine both invariants for complete bipartite graphs $K_{m,n}$ and show that for some choices of m and n the two parameters may be quite far apart. In particular, this gives the first examples of graphs on which the maximum likelihood threshold and the generic completion rank do not agree.

Grigoriy Blekherman
greg@math.gatech.edu

Rainer Sinn
rainer.sinn@fu-berlin.de

[1] School of Mathematics, Georgia Tech, Atlanta, GA 30332, USA

[2] Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin 14195, Germany

Introduction

Gaussian graphical models are an important class of statistical models widely used in system biology and bioinformatics [13,22,28]. In such applications, one is often given data from few observations compared to the number of random variables. The maximum likelihood estimator is computed to describe the interaction between different variables, e.g. genes. It is well known that the maximum likelihood estimator exists with probability one if the number of observations is at least as large as the number of variables. However, with fewer observations the maximum likelihood estimator may fail to exist with high probability. Therefore, it is an important problem to find the minimal number of observations such that the sample covariance matrix has a maximum likelihood estimator with probability one in terms of the graph G underlying the Gaussian graphical model. This number was studied in [8,29] and following [18] we call it the *maximum likelihood threshold* of G, denoted $\text{mlt}(G)$. See [25] and [30] for an introduction to Gaussian graphical models.

Gaussian graphical models were introduced by Dempster in [11]. In his seminal paper, he showed that existence of the maximum likelihood estimator is equivalent to a positive definite matrix completion problem. Using this reformulation, Buhl proved upper and lower bounds on the maximum likelihood threshold in [8]. Uhler in [29] introduced an algebraic relaxation of the maximum likelihood threshold which we will call the *generic completion rank* of G, which provides an upper bound on the maximum likelihood threshold. Gross and Sullivant made a connection between generic completion rank and combinatorial rigidity theory to prove upper bounds on the generic completion rank, and by extension on the maximum likelihood threshold of G. However, the work of Uhler and Gross–Sullivant left open the question of whether the maximal likelihood threshold and the generic completion rank are different for any graph G, as they agree in the small examples where both could be computed.

In this paper, we show that the maximal likelihood threshold and the generic completion rank agree under *clique sums* of graphs. Therefore, large classes of graphs on which these two parameters agree can be built using known cases of equality, see Corollary 1.16. However, we also compute both the maximal likelihood threshold and the generic completion rank of complete bipartite graphs $K_{m,n}$. It turns out that these quantities can be quite different in this case. This provides the first examples of graphs where the two parameters do not agree. This phenomenon is also interesting for practical purposes, as it shows that maximum likelihood estimator may exist with probability one with a much smaller number of observations than suggested by the generic completion rank.

We make several reinterpretations of the maximal likelihood threshold and the generic completion rank which we think are interesting in themselves, and we relate generic completion rank of a symmetric matrix to the non-symmetric case. This shows that the generic completion rank is an important parameter in itself, in need of further study and understanding. We now review previous work and our results in detail.

Main Results We write $\mathcal{S}^n$ for the vector space of real symmetric $n \times n$ matrices and $\mathcal{S}^n_+$ for the cone of positive semidefinite symmetric matrices. Let $G = (V, E)$ be a simple graph on n vertices labeled by the integers $[n] = \{1, 2, \ldots, n\}$. Given G, we

have a projection π_G from the real vector space $\mathcal{S}^n$ to the real vector space $\mathbb{R}^n \oplus \mathbb{R}^E$, where the coordinates in the first summand are indexed by vertices of G and in the second summand by edges of G. It is defined as

$$\pi_G(a_{ij}) = (a_{ii} : i \in [n]) \oplus (a_{ij} : \{i, j\} \in E).$$

The image of this projection is a partially specified symmetric matrix and we call it a *G-partial matrix*. The existence of a unique maximum likelihood estimator of a G-partial matrix $\pi_G(A)$ in a Gaussian graphical model is equivalent to the existence of a positive definite matrix completion of $\pi_G(A)$ by a seminal result of Dempster [11]. Concretely, given a sample $X_1, X_2, \ldots, X_r \in \mathbb{R}^n$ drawn from a multivariate Gaussian distribution of the Gaussian graphical model corresponding to G, we compute the sample covariance matrix $A = \sum_{i=1}^r X_i X_i^T$, which is the sum of the rank-1 matrices given by the samples (here, we assume that the mean of the distribution is 0 for simplicity). The maximum likelihood estimator exists and is unique if and only if $\pi_G(A)$ has a positive definite completion [11].

Therefore the question of computing the maximal likelihood threshold is equivalent to the following matrix completion problem:

What is the minimal number r such that for a generic positive semidefinite matrix A of rank r, the projection $\pi_G(A)$ has a positive definite completion?

Buhl [8] proved bounds on $\text{mlt}(G)$ in terms of well-known graph parameters of G. Let $\omega(G)$ be the clique number of G (i.e., the size of the vertex-set in a largest complete subgraph of G, see [12, Sect. 5.5]) and $\tau(G)$ be the tree-width of G (i.e., the smallest clique number minus 1 of a chordal cover of G, see [12, Sect. 12.3]). Then we have the following:

$$\omega(G) \leq \text{mlt}(G) \leq \tau(G) + 1.$$

However, the clique number and the tree-width of a graph may be far apart and these bounds are often far from sharp. In [29], Uhler introduced an algebraic relaxation of maximum likelihood threshold:

The *generic completion rank* of the graph G is the smallest integer r such that $\dim(\pi(V_r^n)) = n + \#E$ where $V_r^n \subset \mathcal{S}^n$ denotes the variety of symmetric matrices of rank at most r, see Definition 1.1. The generic completion rank of G is called the rank of G in [18].

Uhler [29] showed that

$$\text{mlt}(G) \leq \text{gcr}(G),$$

and, using this inequality, was able to compute the maximal likelihood threshold in some examples.

Gross and Sullivant [18] made a fruitful connection between generic completion rank of G and combinatorial rigidity theory to prove some very interesting results. For instance, they showed that the generic completion rank (and therefore the maximum likelihood threshold) of any planar graph is at most 4. In all of the examples where

both parameters could be computed, there is equality between the maximal likelihood threshold and the generic completion rank. For non-symmetric completion problems considered in Sect. 3, Bernstein recently gave a combinatorial classification of bipartite graphs whose generic completion rank is two using tropical geometry tools [4].

We first show that maximal likelihood threshold and generic completion rank behave in the same way under clique sums. Recall that given two simple graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ such that the induced subgraphs on $V_1 \cap V_2$ are cliques in G_1 and G_2, respectively, their *clique sum* is the graph $(V_1 \cup V_2, E_1 \cup E_2)$. We denote it by $G_1 \boxtimes G_2$. We show the following theorem:

Theorem (Theorems 1.13 and 1.15). *For any simple graphs G_1 and G_2*

$$\mathrm{mlt}(G_1 \boxtimes G_2) = \max\{\mathrm{mlt}(G_1), \mathrm{mlt}(G_2)\}$$
$$and \quad \mathrm{gcr}(G_1 \boxtimes G_2) = \max\{\mathrm{gcr}(G_1), \mathrm{gcr}(G_2)\}.$$

Another notion related to the generic completion rank, called the Gaussian rank of a graph, was introduced by Ben-David in [3]. The same statement for clique sums also holds for this notion, see [3, Prop. 3.4].

Using clique sums, we can build large families of graphs on which the generic completion rank and maximum likelihood threshold coincide, see Corollary 1.16 for a characterization in terms of induced subgraphs. On the other hand, we compute the generic completion rank and the maximum likelihood threshold of complete bipartite graphs. A result related to the computation of the generic completion rank was obtained in the rigidity literature [7,27]. We explain a relation between the results in more detail in Sect. 2, after Theorem 2.1.

Theorem *Let $n \geq m \geq 2$ and set M to be the smallest k such that $\binom{k+1}{2} \geq m + n$.*

(1) *The generic completion rank of $K_{m,n}$ is $m + 1$ if $n > \binom{m}{2}$. It is m if $n \leq \binom{m}{2}$ (Theorem 2.1).*

(2) *The maximum likelihood threshold of the complete bipartite graph $K_{m,n}$ is the minimum between M and $m + 1$ (Theorem 2.7).*

Note that the generic completion rank of $K_{m,n}$ is m or $m + 1$, depending on the size of the larger part n compared to m, and therefore, this is an upper bound on the maximum likelihood threshold. The above minimum is M for $n \leq \binom{m}{2}$ and $m + 1$ for $n > \binom{m}{2}$. Therefore we have the following corollary:

Corollary (Corollary 2.11).

(a) *The maximum likelihood threshold of $K_{5,5}$ is 4, whereas the generic completion rank is 5.*

(b) *Asymptotically, the maximum likelihood threshold of $K_{m,m}$ is $2\sqrt{m}$.*

(c) *The maximum likelihood threshold of $K_{m,n}$ is equal to the generic completion rank of $K_{m,n}$ for $m = 2, 3, 4$ and whenever $n > \binom{m}{2}$.*

The case $m = 2$ in (c) was known before, see [29, Prop. 4.2].

The case of complete bipartite graphs also shows that the maximum likelihood threshold can be far from the Gaussian rank of a graph because the Gaussian rank of $K_{m,n}$ is its treewidth, i.e., $m + 1$, by a result of Ben-David [3, Thm. 1.1].

In the end of this paper, we relate the generic completion rank of bipartite graphs to the generic completion rank for the general matrix completion problem: Given a partially specified $m \times n$ matrix, we form a bipartite graph on the row and column indices by adding the edge $\{i, j\}$ if and only if the (i, j)th entry of the partial matrix is given. We show that the generic completion rank of the graph obtained this way is (up to an additive constant of at most 1) equal to the generic completion rank in the general case, see Proposition 3.1. We take this as an indication that the generic completion rank of bipartite graphs is an important graph parameter.

We end this section with an open question of whether the maximum likelihood threshold can be used to provide an upper bound for the generic completion rank. As explained above for complete bipartite graphs $K_{m,m}$, the generic completion rank is a quadratic function of maximum likelihood threshold. But it is not known whether there exist graphs with bounded maximum likelihood threshold and arbitrarily high generic completion rank.

Question *Does there exist a function* $f : \mathbb{N} \to \mathbb{N}$ *such that for all* $k \in \mathbb{N}$:

$$\mathrm{mlt}(G) \leq k \quad \textit{implies} \quad \mathrm{gcr}(G) \leq f(k).$$

1 Generic completion rank and maximum likelihood threshold

We write $\mathcal{S}^n$ for the vector space of real symmetric $n \times n$ matrices and $\mathcal{S}^n_+$ for the cone of positive semidefinite symmetric matrices. We denote the interior of $\mathcal{S}^n_+$ by $\mathcal{S}^n_{++}$; $\mathcal{S}^n_{++}$ is the cone of positive definite matrices. Let $G = (V, E)$ be a simple graph on n vertices labeled by the integers $[n] = \{1, 2, \ldots, n\}$. Given G, we have a projection π_G from the real vector space $\mathcal{S}^n$ of real symmetric $n \times n$ matrices to the real vector space $\mathbb{R}^n \oplus \mathbb{R}^E$, where the coordinates in the first summand are indexed by vertices of G and in the second summand by edges of G. It is defined as

$$\pi_G(a_{ij}) = (a_{ii} : i \in [n]) \oplus (a_{ij} : \{i, j\} \in E).$$

We think of a vector in the image of this projection as a partially specified symmetric matrix and call it a *G-partial matrix*. We are interested in the positive semidefinite matrix completion problem, i.e., we would like to describe the image of the cone of positive semidefinite real symmetric matrices $\mathcal{S}^n_+$ under the coordinate projection π_G, which is the convex cone of G-partial matrices with a positive semidefinite completion. We write Σ_G for $\pi_G(\mathcal{S}^n_+)$.

Definition 1.1 The *generic completion rank* of the graph $G = ([n], E)$ is the smallest integer r such that the projection π_G restricted to the variety of matrices of rank at most r is dominant. In other words,

$$\dim(\pi(V_r^n)) = n + \#E,$$

where $V_r^n \subset \mathcal{S}^n$ denotes the variety of symmetric matrices of rank at most r.

The generic completion rank of G is called the rank of G in [18]. By Chevalley's Theorem [20, Exer. II.3.19], the set $\pi_G(V_r^n)$ is constructible in the Zariski topology. This implies that almost every G-partial matrix will have a completion with complex entries of rank equal to the generic completion rank of G, which motivates the name.

We will give equivalent definitions from an algebro-geometric point of view. We fix the usual trace inner product on the real vector space of symmetric matrices, i.e., $\langle A, B \rangle = \mathrm{tr}(AB)$ for $A, B \in \mathcal{S}^n$. The orthogonal complement of the kernel of π_G is the coordinate subspace of matrices whose (i, j)th entry is 0 whenever $\{i, j\}$ is not an edge. We denote this linear space by L_G.

We interpret the kernel of the projection π_G as a linear space of quadratic forms by the one-to-one correspondence $Q_M = (x_1, \ldots, x_n)M(x_1, \ldots, x_n)^t$; it is the span of the quadratic forms $x_i x_j$ for all non-edges $\{i, j\}$ of the graph. We write I_G for the ideal generated by $\ker(\pi_G)$. This is a square-free monomial ideal and the Stanley–Reisner ideal of the clique complex of the graph. It defines an algebraic set $X_G = \mathcal{V}(I_G) \subset \mathbb{P}^{n-1}$, which is a subspace arrangement. In fact, X_G is the union of all subspaces $\mathbb{P}(\mathrm{span}\{e_i : i \in K\})$, where $K \subset G$ is a clique and e_i are the standard basis vectors in $\mathbb{R}^n$. We write R for the homogeneous coordinate ring $\mathbb{R}[x_1, \ldots, x_n]/I_G$ of X_G, which is graded by the total degree grading. We write R_i for the homogeneous part of degree i. In this geometric setup, the convex cone $\pi_G(\mathcal{S}_+^n)$ is a subset of R_2, namely the cone of sums of squares on X_G, hence the notation Σ_G, see [6, Sect. 6.1].

In this paper, a *linear series* on X_G is a linear subspace of R_1. Using these notations, we give two equivalent definitions of the generic completion rank. We write $\langle S \rangle$ for the ideal generated by a subset $S \subset R$ and $\langle S \rangle_i$ for the homogeneous part of degree i of a homogeneous ideal.

Proposition 1.2 *The integers defined in the following statements are equal to the generic completion rank of G.*

(1) *The smallest k such that there exists a linear series $W \subset R_1$ of dimension k on X_G with the property that $\langle W \rangle_2 = R_2$.*

(2) *The smallest k such that there exists a linear series $W \subset R_1$ of dimension k on X_G which is not contained in the kernel of a matrix in the orthogonal complement of $\ker(\pi_G)$.*

Proof The generic completion rank is equal to the integer in (1) by Terracini's Lemma, see Flenner–O'Carroll–Vogel [15, Prop. 4.3.2]. Indeed, the equality $\dim(\pi_G(V_r^n)) = n + \#E$ holds if and only if the differential at a generic point on V_r^n is surjective by Generic Smoothness [20, Corr. III.10.7 and Proposition III.10.4]. Terracini's Lemma says that the tangent space to V_r^n at a generic point $M = \ell_1^2 + \cdots + \ell_r^2$ is the degree 2 part of the homogeneous ideal generated by $\ell_1, \ldots, \ell_r$. So the differential is surjective if and only if $T_M V_r^n + \ker(\pi_G) = \mathcal{S}^n$, which is in turn equivalent to $\langle \ell_1, \ldots, \ell_r \rangle_2 = R_2 \cong \mathcal{S}^n / \ker(\pi_G)$.

We now prove that the two integers defined in (1) and (2) are equal. By duality in linear algebra, we have $\langle W \rangle_2 = R_2$ for a linear series $W \subset R_1$ if and only if there is no linear functional $\ell \in R_2^*$ that vanishes on $\langle W \rangle_2$. To a linear functional $\ell \in R_2^*$, we associate the quadratic form $Q_\ell : R_1 \to \mathbb{R}$, $Q_\ell(f) = \ell(f^2)$; the representing matrix of Q_ℓ with respect to the monomial basis is the moment matrix or middle catalecticant

of ℓ. Then ℓ vanishes on $\langle W \rangle_2$ if and only if W is in the kernel of the moment matrix of ℓ. With the trace inner product, we identify R_2^* with the orthogonal complement of $\ker(\pi_G)$, i.e., $R_2^* \cong L_G$. So there exists a linear series $W \subset R_1$ of dimension k such that $\langle W \rangle_2 = R_2$ if and only if there is no matrix in the orthogonal complement of $\ker(\pi_G)$ that contains W in its kernel. $\qquad\square$

Remark 1.3 In the setup of the preceding Proposition 1.2, the existence of a linear series $W \subset R_1$ of dimension k on X_G with $\langle W \rangle_2 = R_2$ implies that the same holds for a generic linear series $W' \subset R_1$ of dimension k. To see this, choose a basis $f_1, \ldots, f_k$ of W and a basis $x_1, x_2, \ldots, x_n$ of R_1. Then $\langle W \rangle_2$ is the span of the elements $x_i f_j$ in R_2. If we write these elements as vectors with respect to a fixed basis of R_2, this is a statement about the rank of a matrix. Therefore, upper semi-continuity of the rank of a matrix shows that $\dim \langle W' \rangle_2 < \dim(R_2)$ imposes algebraic conditions on the entries of basis vectors $f_1', f_2', \ldots, f_k'$ of the linear series W' coming from the vanishing of the appropriate minors.

Using these equivalent characterizations of the generic completion rank, we can recover e.g. [18, Thm. 3.2] by a simple dimension count.

Theorem 1.4 *Let G be a simple graph on n vertices and suppose that the generic completion rank of G is r. Then G has at most $n(r-1) - \binom{r}{2}$ edges.*

Proof The number of edges of G can be expressed in terms of the kernel of π_G, namely $n + \#E + \dim(\ker(\pi_G)) = \binom{n+1}{2}$. The fact that the generic completion rank of G is r implies that $\dim(V_r^n) + \dim(\ker(\pi_G)) \geq \dim(\mathcal{S}^n)$. The dimension of the variety of symmetric $n \times n$ matrices of rank at most r is $rn - \binom{r}{2}$, so we get

$$\left[rn - \binom{r}{2} \right] + \left[\binom{n}{2} - \#E \right] \geq \binom{n+1}{2}.$$

The claimed inequality follows by simplification of this one. $\qquad\square$

We now turn to the interior of Σ_G, which is the set of all G-partial matrices that have a positive definite completion, see [9, Prop. 5.5] or [10, Lem. 1.5].

The set of all symmetric matrices A of rank r such that $\pi_G(A) \in \mathrm{int}(\Sigma_G)$ is a semialgebraic subset of the variety V_r^n of matrices of rank at most r. In the following, we say that *almost all* positive semidefinite matrices of rank r map to $\mathrm{int}(\Sigma_G)$ if the set of positive semidefinite matrices of rank r mapping to the boundary of Σ_G is contained in a proper algebraic subvariety of V_r^n. In other words, a positive semidefinite matrix of rank r will map to the interior of Σ_G with probability 1.

Definition 1.5 The *maximum likelihood threshold* of G is the smallest integer r such that for almost all positive semidefinite matrices A of rank r there exists a positive definite matrix P with

$$\pi_G(A) = \pi_G(P).$$

As for the generic completion rank in Proposition 1.2, we will also give equivalent definitions from an algebro-geometric point of view which look very similar.

Proposition 1.6 *The integers defined in the following statements are equal to the maximum likelihood threshold of G.*

(1) *The smallest k such that the vector space $\langle W \rangle_2 + \ker(\pi_G) \subset \mathcal{S}^n$ contains a positive definite matrix for a generic linear series $W \subset R_1$ of dimension k.*

(2) *The smallest k such that a generic linear series $W \subset R_1$ of dimension k is not in the kernel of a positive semidefinite matrix in the orthogonal complement of $\ker(\pi_G)$.*

Proof As before, we first show that our definition of the maximum likelihood threshold is equivalent to (1). First, let $A = \ell_1^2 + \cdots + \ell_r^2$ be a generic positive semidefinite matrix of rank r such that $\pi_G(A) = \pi_G(P)$ for a positive definite matrix P. Set $W = \text{span}\{\ell_1, \ldots, \ell_r\} \subset R_1$. Then $P \in \langle W \rangle_2 + \ker(\pi_G)$. On the other hand, let $W = \text{span}\{\ell_1, \ldots, \ell_r\} \subset R_1$ be a linear series such that $\langle W \rangle_2 + \ker(\pi_G)$ contains a positive definite matrix. Then the G-partial matrix $\pi_G(A)$ for $A = \ell_1^2 + \cdots + \ell_r^2$ has a positive definite completion. We need to show that this holds for generic matrices. The choice of basis $\{\ell_1, \ldots, \ell_r\}$ of W corresponds to the choice of a point $\ell_1^2 + \cdots + \ell_r^2$ in the relative interior of a face of $\mathcal{S}_+^n$, so that generic linear series $W \subset R_1$ of dimension r correspond to generic positive semidefinite matrices of rank r, see [2, Sect. II.12].

We now show that the integers in (1) and (2) are equal. By the separation theorem in convex geometry and the fact that the cone of positive semidefinite matrices is self-dual with respect to the trace inner product [2, Exer. IV.5.3.2], we know that the vector space $\langle W \rangle_2 + \ker(\pi_G)$ contains a positive definite matrix if and only if there is no positive semidefinite matrix in the orthogonal complement L_G that contains W in its kernel. This shows that the integers in (1) and (2) are the same.   □

The above reformulations of the generic completion rank and the maximum likelihood threshold of a graph make it clear that the generic completion rank is the natural algebraic relaxation of the maximum likelihood threshold, a fact first observed by Uhler [29].

Corollary 1.7 *The maximum likelihood threshold of a graph is bounded above by its generic completion rank.*

Proof This follows from the equivalent characterizations of the two invariants in Propositions 1.2(1) and 1.6(1).   □

We will study how generic completion rank and maximum likelihood threshold behave under graph operations. Both invariants are non-increasing with respect to edge and vertex deletion (this is straightforward for the generic completion rank; for the maximum likelihood threshold, the reason is that linear projections are open maps, see the proof of Theorem 1.15 for an argument). However, unlike a related parameter of Gram dimension introduced in [24], the properties $\text{mlt}(G) \leq k$ and $\text{gcr}(G) \leq k$ are not minor-closed as the following example shows.

Example 1.8 Start with the complete graph K_4 on four vertices and divide every edge in such a way that we get the graph G in Figure 1.

By construction, K_4 is a minor of G. The generic completion rank and maximum likelihood threshold of K_4 are 4. By inductively deleting vertices of degree 2, we end

Fig. 1 The graph in
Example 1.8
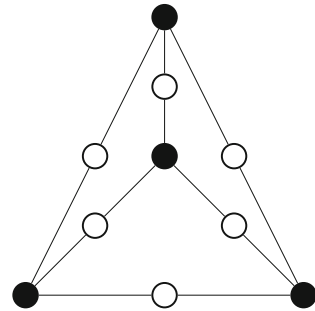
up with an empty 3-core, which shows that the generic completion rank of G is at most 3 by [18, Prop. 3.5] (see also [18, Thm. 3.7]).

From the geometric point of view, we can understand the clique sum of graphs well.

Definition 1.9 Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two simple graphs such that the induced subgraphs on $V_1 \cap V_2$ are cliques in G_1 and G_2, respectively. Their *clique sum*, denoted by $G_1 \boxtimes G_2$, is the graph $(V_1 \cup V_2, E_1 \cup E_2)$.

The geometric operation corresponding to the clique sum of graphs is referred to as linear join in the algebraic geometry literature, see [14]. We introduce this language here because we will later use results from [6] which are stated in this language.

Definition 1.10 Let X_1 and X_2 be two subspace arrangements in $\mathbb{P}^{n-1}$. We say that $X = X_1 \cup X_2$ is *linearly joined* if $X_1 \cap X_2 = \langle X_1 \rangle \cap \langle X_2 \rangle$, where $\langle X_i \rangle$ denotes the span of X_i in $\mathbb{P}^{n-1}$.

Proposition 1.11 *Suppose G is the clique sum of two subgraphs G_1 and G_2 along a clique $K = G_1 \cap G_2$. Then $X_G = X_{G_1} \cup X_{G_2}$ is linearly joined along the coordinate subspace $X_K = \langle \{e_i : i \in K\} \rangle$.*

Proof The span of X_{G_i} is the coordinate subspace spanned by the coordinates indexed by vertices in G_i. Therefore, the two coordinate subspaces $\langle X_{G_1} \rangle$ and $\langle X_{G_2} \rangle$ intersect along the coordinate subspace spanned by the coordinates indexed by vertices in $K = G_1 \cap G_2$. □

We will show that the generic completion rank is well-behaved under clique sums. We need the following lemma in the proof of Theorem 1.13.

Lemma 1.12 *Let M be an $n \times n$ symmetric matrix of full rank with complex entries. Let $a_0, a_1, \ldots, a_n, b_1, b_2, \ldots, b_{n+1}$ be complex numbers and consider the matrix*

$$
P(t) = \begin{pmatrix}
a_0 & a_1 & \ldots & a_n & t \\
a_1 & & & & b_1 \\
\vdots & & M & & \vdots \\
a_n & & & & b_n \\
t & b_1 & \ldots & b_n & b_{n+1}
\end{pmatrix},
$$

which is symmetric of size $(n+2) \times (n+2)$.

(a) *There exists a value $t_0 \in \mathbb{C}$ such that the matrix $P(t_0)$ has rank at most $n + 1$.*
(b) *If the upper left $(n + 1) \times (n + 1)$ block has full rank $n + 1$ and the bottom right $(n + 1) \times (n + 1)$ block has rank n, then there is a unique value $t_0 \in \mathbb{C}$ such that $P(t_0)$ has rank $n + 1$. This value is determined by the linear relation among the rows (or equivalently, the columns) of the bottom right block.*
(c) *If the upper left $(n + 1) \times (n + 1)$ as well as the bottom right $(n + 1) \times (n + 1)$ block have rank n then there exists $t_0 \in \mathbb{C}$ such that $P(t_0)$ has rank n. The entry t_0 is determined by the linear relations among the rows of the bottom right block.*

Proof The coefficient at t^2 in the determinant of $P(t)$ is the determinant of M, which is non-zero. Any zero of the determinant is a value t_0 satisfying (a). For part (b), we use Schur's determinantal formula [31, Thm. 1.1]. We have

$$\det(P(t)) = \det(A) \det\big(b_{n+1} - (t, b_1, \ldots, b_n) A^{-1} (t, b_1, \ldots, b_n)^t\big)$$
$$= \det(A)\big(b_{n+1} - (t, b_1, \ldots, b_n) A^{-1} (t, b_1, \ldots, b_n)^t\big),$$

where A is the upper left $(n + 1) \times (n + 1)$ block of $P(t)$, which is invertible by assumption. We show that this polynomial in t is a square. Essentially, this is because b_{n+1} is equal to $\underline{b}^t M^{-1} \underline{b}$, which follows from Schur's determinantal formula applied to the bottom right $(n + 1) \times (n + 1)$ block, where $\underline{b} = (b_1, \ldots, b_n)^t$. We have

$$A^{-1} = \begin{pmatrix} 1 & 0 \\ -M^{-1}\underline{a} & I_n \end{pmatrix} \begin{pmatrix} a_0 - \underline{a}^t M^{-1}\underline{a} & 0 \\ 0 & M^{-1} \end{pmatrix} \begin{pmatrix} 1 & -\underline{a}^t M^{-1} \\ 0 & I_n \end{pmatrix}$$

by [31, Thm. 1.2] (see the proof). Combining these results using the symmetry of $P(t)$ shows that the determinant is indeed a square, namely

$$\det(P(t)) = (a_0 - \underline{a}^t M^{-1}\underline{a})(t - \underline{b}^t M^{-1}\underline{a})^2,$$

where $\underline{a} = (a_1, \ldots, a_n)^t$. The zero of the determinant is the entry determined by the linear relations among the columns of the bottom right $(n + 1) \times (n + 1)$ block. To prove part (c), write $v_1, \ldots, v_n$ for the rows of the matrix M. Since the upper left $(n+1) \times (n+1)$ block of P has rank n, there are $u_1, \ldots, u_n$ such that $(a_0, a_1, \ldots, a_n) = u_1(a_1, v_1) + u_2(a_2, v_2) + \cdots + u_n(a_n, v_n)$. Arguing similarly for the bottom right $(n + 1) \times (n + 1)$ block, we find $u_1', u_2', \ldots, u_n'$ such that $(b_1, \ldots, b_n, b_{n+1}) = u_1'(v_1, b_1) + \cdots + u_n'(v_n, b_n)$. Setting the missing entry to be $u_1 b_1 + u_2 b_2 + \cdots + u_n b_n$ proves the claim. Indeed, it is enough to show that $u_1 b_1 + u_2 b_2 + \cdots + u_n b_n = u_1' a_1 + u_2' b_2 + \cdots + u_n' b_n$. This follows from symmetry of M because

$$(u_1, \ldots, u_n, 0)(b_1, \ldots, b_{n+1})^t$$
$$= (u_1, \ldots, u_n, 0)\big(u_1'(v_1, b_1) + \cdots + u_n'(v_n, b_n)\big)^t$$
$$= (0, u_1', \ldots, u_n')\big(u_1(a_1, v_1) + \cdots + u_n(a_n, v_n)\big)^t$$
$$= (0, u_1', \ldots, u_n')(a_0, a_1, \ldots, a_n)^t.$$

$\square$

Theorem 1.13 *Let G be the clique sum of two subgraphs G_1 and G_2. Then the generic completion rank of G is equal to the maximum of the generic completion ranks of G_1 and G_2.*

This statement also follows from known results in rigidity theory [26, Thm. 2.1] using the connection to stress-free embeddings of graphs explained in [18, Sect. 2]. We will give an elementary proof, using algorithmic linear algebra for completeness.

Proof of Theorem 1.13 Let G be the clique sum of G_1 and G_2 along a clique $K \subset G$ of size k. We prove the claim by taking the geometric point of view. The generic completion rank of G is the smallest r such that a generic matrix in $\mathbb{C}[X_G]_2 = (\mathbb{C}[x_1, \ldots, x_n]/I_G)_2$ can be represented by a quadratic form in $\mathbb{C}[x_1, \ldots, x_n]_2$ of rank r. After relabeling the vertices of G, if necessary, we can assume that an extension of a quadratic form in $\mathbb{C}[X_G]_2$ is a block matrix

$$\begin{pmatrix} \begin{array}{cc} \boxed{\begin{matrix} A_1 & \\ & \boxed{C} \end{matrix}} & ? \\ ? & \boxed{\begin{matrix} \\ A_2 \end{matrix}} \end{array} \end{pmatrix}$$

where the upper left block is the restriction to X_{G_1} and the lower right block is the restriction to X_{G_2}. Note that they have a common $k \times k$ submatrix, denoted C, which is the restriction to $X_{G_1} \cap X_{G_2} = \langle e_i : i \in K \rangle$. The entries in the top right and lower left are undetermined in $\mathbb{C}[X_G]_2$ and choices of all unknown unspecified entries in this matrix give an extension to $\mathbb{C}[x_1, \ldots, x_n]_2$.

If the entire matrix has a completion of rank r, then so do the top left block and bottom right block. This shows that the generic completion rank of G is at least the maximum of the generic completion ranks of G_1 and G_2.

For the reverse inequality, let the generic completion ranks of G_1 and G_2 be r_1 and r_2, respectively, and assume $r_1 \geq r_2$. We will inductively construct a completion of rank r_1 using the preceding Lemma 1.12. First, given generic entries, we can complete the blocks A_1 and A_2 to matrices of rank r_1 and r_2, respectively. Further, we can assume by genericity that the bottom right $r_1 \times r_1$ block of A_1 has full rank and the top left $r_2 \times r_2$ block of A_2 has full rank. If $r_2 > k = \text{rank}(C)$, we grow A_2 by one row in the following way (compare Figure 2): Add generic entries (for the stars) in the new row until we are at the r_2th column of A_2 (the triangle). Choose this entry so that the determinant of the $(r_2 + 1) \times (r_2 + 1)$ block formed with entries from A_1 in the new row is 0 (Lemma 1.12 (a)), i.e., the last column of this block is a linear combination of the first r_2 columns. By Lemma 1.12 (b), we can fill in the entire new row by the representation of every column of A_2 as a linear combination of the first r_2. In this way, we get an extension A_2' of A_2 by one row such that the first r_2 columns of A_2' are a basis of the column span of A_2' and the top left $r_2 \times r_2$ block has full rank. This process enlarges the common block C of A_1 and A_2 and grows its rank by 1. So we can inductively add rows to A_2 until $r_2 = k$.

If $r_1 > r_2 = k$, we add one row to A_2 similarly as before: Fill in each entry of the row according to the linear relation of the columns by Lemma 1.12 (b). The obtained

Fig. 2 Visualization of the growing step in the proof of Theorem 1.13
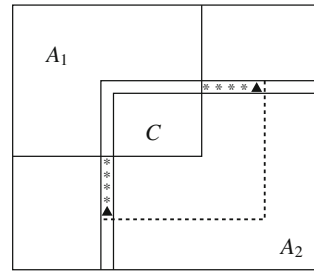
extension A_2' of A_2 will have rank $r_2 + 1$, because the first $r_2 + 1$ columns of A_2' are now a basis of the column span of A_2'.

So we have now reduced to the case that $r_1 = r_2 = k$. Now we fill in the missing entries in the top right block by the linear relations among the rows of A_1, which gives the same as using the linear relations among the rows of A_2 in the bottom left block by Lemma 1.12 (c). In the end, we get a completion of rank r_1, which proves the claim. $\square$

We now examine the behavior of the maximum likelihood threshold under clique sums.

Lemma 1.14 *Let G be the clique sum of G_1 and G_2. Then a quadratic form $f \in \mathbb{R}[X_G]_2$ lies in the interior of Σ_G if and only if its restrictions to X_{G_1} and X_{G_2} lie in the interior of Σ_{G_1} and Σ_{G_2}, respectively.*

Proof Since G is a clique sum of graphs, $X_G = X_{G_1} \cup X_{G_2}$ is linearly joined. By [6, Corr. 5.4], we have $\Sigma_G^\vee = \text{conv}(\Sigma_{G_1}^\vee \cup \Sigma_{G_2}^\vee)$, where both cones on the right hand side are faces of $\Sigma_G^\vee$. In particular, the union of extreme rays of $\Sigma_G^\vee$ is the union of the extreme rays of $\Sigma_{G_1}^\vee$ and $\Sigma_{G_2}^\vee$. By the separation theorem in convex geometry, a point x is in the interior of Σ_G if and only if for every extreme ray $\mathbb{R}_+\ell$ of $\Sigma_G^\vee$ we have $\ell(x) > 0$. This proves the claim. $\square$

Theorem 1.15 *Let G be the clique sum of G_1 and G_2. The maximum likelihood threshold of G is the maximum of the maximum likelihood thresholds of G_1 and G_2.*

Proof We assume, after relabeling of the vertices, if necessary, that $G_1 = ([p], E_1)$ and $G_2 = ([q] + r, E_2)$, so that $\dim \langle X_{G_1} \rangle = p - 1$ and $\dim \langle X_{G_2} \rangle = q - 1$ (here $[q] + r$ is the set of integers $\{r + 1, r + 2, \ldots, r + q\}$ and $|G_1 \cap G_2| = p - r$). Then the composition of the projection $\pi_G : \mathcal{S}^n \to \mathbb{R}[X_G]_2$ with the restriction map $\text{res}_1 : \mathbb{R}[X_G]_2 \to \mathbb{R}[X_{G_1}]_2$ is the projection $\pi_1 : \mathcal{S}^p \to \mathbb{R}[X_{G_1}]_2$ of the upper left $p \times p$ block of the $n \times n$ matrix. The analogous statement for the lower right $q \times q$ block holds for $\pi_2 : \mathcal{S}^q \to \mathbb{R}[X_{G_2}]_2$.

First, we show $\text{mlt}(G) \leq \max\{\text{mlt}(G_1), \text{mlt}(G_2)\}$. Let $r = \max\{\text{mlt}(G_1), \text{mlt}(G_2)\}$ and let A be a generic positive semidefinite rank r matrix in $\mathcal{S}^n$. Then the upper left $p \times p$ block A_1 and the lower right $q \times q$ block A_2 of A are generic matrices of rank $\min\{r, p\}$ and $\min\{r, q\}$, respectively, with the property that the restriction of $\pi(A)$ to X_{G_1} and X_{G_2} are the projections $\pi_1(A_1)$ and $\pi_2(A_2)$ of the blocks. By definition of the maximum likelihood threshold, we know that $\pi_1(A_1)$ is an interior point of

Σ_{G_1} and $\pi_2(A_2)$ is an interior point of Σ_{G_1}. So $\pi(A)$ is an interior point of Σ_G by Lemma 1.14, which shows the desired inequality.

The reverse inequality $\mathrm{mlt}(G_1) \leq \mathrm{mlt}(G)$ is straightforward since G_1 is a subgraph of G. Indeed, if A is a generic positive semidefinite matrix of rank $r \geq \mathrm{mlt}(G)$, then $\pi_G(A)$ is an interior point of Σ_G. Since projections are open and Σ_{G_1} is the projection of Σ_G, the G_1-partial matrix $\pi_{G_1}(A)$ is an interior point of Σ_{G_1}. $\qquad\square$

Corollary 1.16 *The generic completion rank and maximum likelihood threshold are equal for any graph that does not contain a wheel W_k for $k \geq 5$ or a splitting of a wheel W_k for $k \geq 4$ as an induced subgraph.*

The wheel W_k is a graph on k vertices which is composed of a cycle C_{k-1} on $k-1$ nodes together with an additional vertex adjacent to every vertex of C_{k-1}.

Proof By a result of Johnson and McKee [21], these graphs are clique sums of chordal graphs and series–parallel graphs. Moreover, results of [8] and [29] imply that generic completion rank and maximum likelihood threshold also agree on series–parallel graphs, which are graphs of treewidth at most 2. Therefore, for clique sums of chordal and series–parallel graphs, generic completion rank and maximum likelihood threshold are equal by Theorems 1.13 and 1.15. $\qquad\square$

We will see in Sect. 2 that generic completion rank and maximum likelihood threshold are not equal in general.

The sparsity order of a graph G is the largest rank of an extreme ray of $\Sigma_G^\vee$, see [1] or [23]. The positive semidefinite matrix completion theorem [16, Thm. 7] implies that the sparsity order of G is 1 if and only if G is chordal. We now prove an inequality which relates the sparsity order and the generic completion rank of G.

Proposition 1.17 *Let G be a graph on n vertices. Then the sum of the generic completion rank of G and its sparsity order is at most $n + 1$.*

Proof Let k be the sparsity order of G. Then the kernel of an extreme ray of rank k is an $(n-k)$-dimensional linear series in R_1. A result of Blekherman [5, Prop. 4.2] shows that such a linear series will generate a hyperplane in R_2. This means that a generic linear series of dimension $n - k$ generates at least a hyperplane in R_2, so a generic linear series of dimension $n - k + 1$ will certainly generate all of R_2. This implies that the generic completion rank is at most $n - k + 1$ by Proposition 1.2(1), which proves the claimed inequality. $\qquad\square$

This estimate is sharp in some cases.

Example 1.18 Consider a complete bipartite graph $K_{m,n}$ on $m + n$ vertices. Assume $m \leq n$. We will see in Sect. 2 that the generic completion rank of $K_{m,n}$ is m as long as $n \leq \binom{m}{2}$ and $m + 1$ otherwise. The sparsity order of complete bipartite graphs is determined in [17, Thm. 2.1]: it is n as long as $n \leq \binom{m}{2} + 1$ and $\binom{m}{2} + 1$ otherwise. So the sum of the generic completion rank and the sparsity order is always at most $m + n$ and it is equal to $m + n$ only if $n = \binom{m}{2} + 1$.

2 Complete Bipartite Graphs

Let $n \geq m \geq 2$ be positive integers and denote by $K_{m,n}$ the complete bipartite graph on two color classes of size m and n, respectively. We assume $m \geq 2$ so that $K_{m,n}$ is not a tree. In this section, we will first determine the generic completion rank of $K_{m,n}$ and then its maximum likelihood threshold.

Theorem 2.1 *Let $n \geq m \geq 2$. The generic completion rank of $K_{m,n}$ is $m + 1$ if $n > \binom{m}{2}$. It is m if $n \leq \binom{m}{2}$.*

In terms of rigidity theory, the theorem determines the smallest dimension d such that a generic embedding of the complete bipartite graph into $\mathbb{R}^d$ is stress-free (again, see [18, Sect. 2] for the connection). Rigidity properties of complete bipartite graphs have been studied from the viewpoint of infinitesimal flexibility in [7] and [27]. This is related because the results in these papers estimate the generic (column) rank of the Jacobian of the *edge function* (see [7, Sect. 4], [27, Thm. 2]), which is the same as the rigidity matrix up to row and column operations [18, Thm. 2.4]. An embedding of the graph is stress-free if the row rank of the Jacobian of the edge function at the point corresponding to the embedding is maximal. However, the results in [7] and [27] do not directly imply our theorem because they make additional assumptions comparing the matrix size to the embedding dimension. We will give a proof of this theorem below after some preparations. The following remark is a well-known fact from graph theory.

Remark 2.2 If $m \geq 2$, the treewidth of $K_{m,n}$ is $m + 1$.

Remark 2.3 The generic completion rank of $K_{m,n}$ is at least m, simply because a $K_{m,n}$-partial matrix is of the form

$$\begin{pmatrix} D_1 & B \\ B^t & D_2 \end{pmatrix},$$

where B is an $m \times n$ matrix and D_1 is an $m \times m$ matrix, D_2 is an $n \times n$ matrix, both of which are only specified along their diagonals. So generically, a completion of a $K_{m,n}$-partial matrix will have rank at least m, which is the generic rank of B.

Proposition 2.4 *Suppose $n > \binom{m}{2}$. Then the generic completion rank of $K_{m,n}$ is at least $m + 1$.*

Proof The generic completion rank of $K_{m,n}$ is at least m by the previous Remark 2.3. It cannot be m by dimension count. Indeed, the dimension of the variety V_m^{m+n} of symmetric matrices of size $m + n$ and rank at most m is $\binom{m+1}{2} + m \cdot n = \binom{m}{2} + m + m \cdot n$. Since $n > \binom{m}{2}$, the dimension $m + n + m \cdot n$ of the target of the projection $\pi_{K_{m,n}} : \mathcal{S}^{m+n} \to \mathbb{R}^{m+n} \oplus \mathbb{R}^{\#E}$ is larger than the dimension of V_m^{m+n}. Therefore, the restriction of the projection $\pi_{K_{m,n}}$ to V_m^{m+n} cannot be dominant. □

Theorem 2.5 *Suppose $m \geq 2$ and $n = \binom{m}{2}$. Then the generic completion rank of $K_{m,n}$ is equal to m.*

Proof The generic completion rank of $K_{m,n}$ is at least m. We show by example that a generic linear series of dimension m will generate R_2, which implies that the generic completion rank is equal to m by Proposition 1.2 (1).

It suffices to find one such m-dimensional linear series $W \subset R_1$. Let W be the row span of $(I_m \quad A^t)$, where A is an $n \times m$ matrix, namely the following: We fix an order of the $n = \binom{m}{2}$ pairs (i, j), where $1 \leq i < j \leq m$. We set the kth row of A to be $e_i + e_j$, where (i, j) is the kth pair in our ordering and e_i and e_j are the ith, resp. jth standard basis vectors of $\mathbb{R}^m$.

If W did not generate R_2, there would be a non-zero matrix $M \in L_{K_{m,n}} = \ker(\pi_{K_{m,n}})^{\perp} \cong R_2^*$ such that $W \subset \ker(M)$. Write

$$M = \begin{pmatrix} D_m & B \\ B^t & D_n \end{pmatrix},$$

where D_m and D_n are diagonal matrices of size $m \times m$, resp. $n \times n$. So the condition $W \subset \ker(M)$ is equivalent to

$$\begin{pmatrix} D_m + BA \\ B^t + D_n A \end{pmatrix} = \begin{pmatrix} D_m & B \\ B^t & D_n \end{pmatrix} \begin{pmatrix} I_m \\ A \end{pmatrix} = 0.$$

This system of equations implies $D_m - A^t D_n A = 0$ by solving the second row for B^t and plugging $B^t = -D_n A$ into the first equation. Given A, this is a system of $m + \binom{m}{2}$ linear equations in the $m + n$ variables, which are the diagonal entries of D_m and D_n. Since the equations are in row echelon form for the entries of D_m, we focus on the off-diagonal equations in the matrix $D_m - A^t D_n A = 0$ as equations in the diagonal entries of D_n. Write $D_n = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $A = (a_{ij})_{1 \leq i \leq n, 1 \leq j \leq m}$. Then these equations are $\sum_{k=1}^{n} \lambda_k a_{ki} a_{kj} = 0$. The rows of the coefficient matrix C of this linear system in the variables $\lambda_1, \lambda_2, \ldots, \lambda_n$ are indexed by pairs (i, j) for $1 \leq i < j \leq m$ and the (i, j)th row is equal to $(a_{1i}a_{1j}, a_{2i}a_{2j}, \ldots, a_{ni}a_{nj})$. We have chosen A in such a way that the determinant of the coefficient matrix C is non-zero because there is exactly one entry 1 in every row and column of the coefficient matrix C. Indeed, in the row of C indexed by the pair (i, j), only the entry $a_{ki}a_{kj}$ is non-zero and equal to 1, where again (i, j) is the kth pair in the ordering. Similarly, in the kth column of C, only the entry $a_{ki}a_{kj}$ is non-zero and equal to 1. This shows that C is a permutation matrix and therefore has full rank.

Therefore the only solution to the above system of linear equations with coefficient matrix C is $D_n = 0$, which implies $D_m = 0$ and $B^t = 0$ by the above matrix equations. So $M = 0$, which implies that this linear series W generates R_2.   □

Proof of Theorem 2.1 The generic completion rank of $K_{m,n}$ is equal to m in case $n = \binom{m}{2}$ by Proposition 2.5. If $m \leq n < \binom{m}{2}$, the generic completion rank only decreases because we are deleting edges. On the other hand, it cannot drop below m by Remark 2.3, so it must be equal to m. If $n > \binom{m}{2}$, then the generic completion rank is larger than m by Proposition 2.4. Since the treewidth of $K_{m,n}$ is $m + 1$ by Remark 2.2, it cannot be larger than $m+1$ by [18, Prop. 1.3], which finishes the proof.   □

We now turn to the maximum likelihood threshold of complete bipartite graphs. It turns out that it is different from the generic completion rank for certain values of $n \geq m \geq 5$. Below, we use the Grassmannian for a dimension count. Here is a short summary of affine coordinates on it, see e.g., [19].

Remark 2.6 Given a k-dimensional linear space $L \subset \mathbb{R}^m$, we choose a basis $v_1, \ldots, v_k$ and put these vectors as the rows of a matrix. After row operations and possibly permutation of the columns, we can write this matrix as $(I_k \; A)$, where A is a $k \times (m-k)$ matrix. The entries of A are affine coordinates of the linear space L on an affine chart of the Grassmannian $\mathrm{Gr}(k, m)$ of k-dimensional linear subspaces of $\mathbb{R}^m$.

Theorem 2.7 *Let $n \geq m \geq 2$ and set M to be the smallest k such that $\binom{k+1}{2} \geq m + n$. The maximum likelihood threshold of the complete bipartite graph $K_{m,n}$ is the minimum between M and $m + 1$.*

Note that the generic completion rank of $K_{m,n}$ is m or $m + 1$, depending on the size of the larger part n compared to m, and therefore, this is an upper bound on the maximum likelihood threshold. The above minimum is M for $n \leq \binom{m}{2}$ and $m + 1$ for $n > \binom{m}{2}$.

We prove the theorem by two dimension counts, which we state separately. The goal is to provide a lower bound and a matching upper bound using Proposition 1.6(2). We write $\Sigma_{m,n} = \pi_{K_{m,n}}(\mathcal{S}_+^{m+n})$ for the cone of sums of squares of linear forms on the line arrangement $X_{K_{m,n}}$ in $\mathbb{P}^{m+n-1}$ defined by $K_{m,n}$ and $\Sigma_{m,n}^\vee$ for its dual cone. For matrices in the orthogonal complement $L_{K_{m,n}}$ of $\ker(\pi_{K_{m,n}})$, we fix the notation

$$M = \begin{pmatrix} D_m & A \\ A^t & D_n \end{pmatrix},$$

where D_m is an $m \times m$ diagonal matrix, D_n is an $n \times n$ diagonal matrix, and A is an $m \times n$ matrix.

Lemma 2.8 *Let $n \geq m \geq 2$ and set M to be the smallest k such that $\binom{k+1}{2} \geq m + n$. The maximum likelihood threshold of the complete bipartite graph $K_{m,n}$ is at least $\min\{M, m + 1\}$.*

Proof Since every extreme ray of $\Sigma_{m,n}^\vee$ has rank at most n by [17, Thm. 2.1], every linear series that is in the kernel of a matrix $M \in \Sigma_{m,n}^\vee$ is in the kernel of a matrix $M' \in \Sigma_{m,n}^\vee$ of rank at most n. We show that we can assume that M is positive semidefinite of rank n and the diagonal entries of D_m and D_n are non-zero. If some of the diagonal entries of D_m or D_n are equal to 0, then M being positive semidefinite implies that the corresponding rows and columns of M have to be 0. In terms of the graph, this corresponds to the deletion of the vertices corresponding to these zero rows and columns. So this reduces to the case of $K_{m',n'}$, where $m' \leq m$ and $n' \leq n$. So we can assume that the diagonal entries of D_m and D_n are non-zero, which implies that the rank of M is n.

With these assumptions, we conclude from the rank additivity of the Schur complement [31, Sect. 0.9] that $D_m - A D_n^{-1} A^t = 0$. A vector $(v, w)^t$, where $v \in \mathbb{R}^m$ and $w \in \mathbb{R}^n$, is in the kernel of M if and only if $w = -D_n^{-1} A^t v$ by the bottom blocks of M.

It is convenient to interpret this equation differently. Set $\tilde{A} = D_n^{-1/2} A^t D_m^{-1/2}$, where $D_m^{-1/2}$ and $D_n^{-1/2}$ denote the inverse of a square root of D_m and D_n, respectively. Then $D_n^{1/2} w = -\tilde{A} D_m^{1/2} v$ and the above equation $D_m - A D_n^{-1} A^t = 0$ shows that $\tilde{A}$ is an isometric embedding of $\mathbb{R}^m$ into $\mathbb{R}^n$, i.e., $\tilde{A}^t \tilde{A} = I_m$.

We show that for $k \leq m$, there is an open ball $B \subset \mathrm{Gr}(k, m+n)$ in the Euclidean topology such that every k-dimensional subspace $W \in B$ with basis $(v_1, w_1)^t, \ldots, (v_k, w_k)^t$, such that the matrix $(v_1, \ldots, v_k)^t$ has rank k, is contained in the kernel of a matrix $M \in \Sigma_{m,n}^\vee$ whenever $\binom{k+1}{2} < m + n$. To prove this claim, we find an isometric embedding $\tilde{A} \colon \mathbb{R}^m \to \mathbb{R}^n$ and diagonal matrices D_m and D_n such that $D_n w_i = \tilde{A} D_m v_i$. Such an isometric embedding exists if and only if $\langle D_m v_i, D_m v_j \rangle = \langle D_n w_i, D_n w_j \rangle$ for all $i, j = 1, \ldots, k$.

The vectors v_i and w_j are given by the basis of $W \subset \mathbb{R}^{m+n}$. We read these equations as a homogeneous linear system whose variables are squares of the $m+n$ entries of the diagonal matrices D_m and D_n. So we have $\binom{k+1}{2}$ conditions in $m+n$ variables. There exists a nontrivial solution if $\binom{k+1}{2} < m+n$. Therefore, W is contained in the kernel of the matrix

$$M = \begin{pmatrix} D_m^2 & -D_m \tilde{A}^t D_n \\ -D_n \tilde{A} D_m & D_n^2 \end{pmatrix}$$

by the above computation using the Schur complement. By the inertia additivity of the Schur complement, see [31, Thm. 1.12], and the above argument for the existence of $\tilde{A}$, M is positive semidefinite if D_m and D_n have real entries. The semialgebraic set of linear spaces in $\mathrm{Gr}(k, m+n)$ such that D_m and D_n have real entries has non-empty interior, which we will prove in the following Lemma 2.9.

This argument shows that for $k \leq m$, the semialgebraic set of k-dimensional subspaces of $\mathbb{R}^{m+n}$ in $\mathrm{Gr}(k, m+n)$ that are contained in the kernel of positive semidefinite matrices M of rank n has non-empty interior whenever $\binom{k+1}{2} < m+n$. In particular, if $n > \binom{m}{2}$, we find an open ball in $\mathrm{Gr}(m, m+n)$ of subspaces that are in the kernel of positive semidefinite matrices. So by Proposition 1.6, the maximum likelihood threshold of $K_{m,n}$ is at least $\min\{k : \binom{k+1}{2} \geq m+n\}$ for $n \leq \binom{m}{2}$ and at least $m+1$ for $n > \binom{m}{2}$. $\qquad\square$

Lemma 2.9 *Let $k \leq m \leq n$ be integers and assume $\binom{k+1}{2} < m+n$. Then there exists an open ball $B \subset \mathrm{Gr}(k, m+n)(\mathbb{R})$ in the Euclidean topology consisting of real subspaces with basis $(v_1, w_1), (v_2, w_2), \ldots, (v_k, w_k)$, $v_i \in \mathbb{R}^m$ and $w_i \in \mathbb{R}^n$, such that the equations $\langle D_m v_i, D_m v_j \rangle = \langle D_n w_i, D_n w_j \rangle$ for all $1 \leq i \leq j \leq k$, where D_m and D_n are diagonal matrices, have a real solution D_m and D_n of full rank.*

Proof As in the proof of the preceding Lemma 2.8, we think of the equations $\langle D_m v_i, D_m v_j \rangle - \langle D_n w_i, D_n w_j \rangle = 0$ as linear equations in the squares of the diagonal entries of D_m and D_n. Let C be the coefficient matrix of this linear system. Its rows are indexed by pairs (i, j) with $1 \leq i \leq j \leq k$ specifying the equation and its columns are indexed by $a = 1, \ldots, m$, the components of the vectors v_i and then $b = 1, \ldots, n$, the components of the vectors w_i. The $(i, j), a$ entry in the left $\binom{k+1}{2} \times m$ block is given by $v_{i,a} v_{j,a}$ and the $(i, j), b$ entry in the right $\binom{k+1}{2} \times n$ block is given by $-w_{i,b} w_{j,b}$.

Generically, the kernel of C will contain a vector with only non-zero entries. Indeed, if every vector in the kernel of C had a zero entry, there would be a position such that all vectors in the kernel have a zero there. For the matrix C, it means that the corresponding column is not contained in the span of the other columns. By symmetry of the columns of C, there would then be a non-empty open set of vectors $(v_1, w_1), \ldots, (v_k, w_k)$ such that the same holds for each column. Since $(\mathbb{A}^{m+n})^k$ is irreducible, the intersection of these non-empty open sets would be non-empty. This would mean that C has rank $m + n$, which is absurd because $\binom{k+1}{2} < m + n$.

In other words, the following map is dominant:

$$
\gamma : \begin{cases} (\mathbb{C}^*)^m \times (\mathbb{C}^*)^n \times Y \to \mathrm{Gr}(k, m+n) \\[2mm] (D_m, D_n, (v_1, \ldots, v_k, w_1, \ldots, w_k)) \mapsto \begin{pmatrix} \left(D_m^{-1} v_1\right)^t & \left(D_n^{-1} w_1\right)^t \\ \vdots & \vdots \\ \left(D_m^{-1} v_k\right)^t & \left(D_n^{-1} w_k\right)^t \end{pmatrix} \end{cases},
$$

where Y is the variety of vectors $(v_1, \ldots, v_k, w_1, \ldots, w_k)$ in $(\mathbb{A}^m)^k \times (\mathbb{A}^n)^k$ such that $\langle v_i, v_j \rangle = \langle w_i, w_j \rangle$ and $v_1, \ldots, v_k$ are linearly independent. For every irreducible component of Y, its real points are Zariski-dense. So, by Generic Smoothness [20, Corr. III.10.7 and Prop. III.10.4], the image of this map restricted to real matrices D_m and D_n and real vectors $(v_1, \ldots, v_k, w_1, \ldots, w_k)$ has full dimension in the whole image of γ as a semi-algebraic set. This is a reformulation of the claim. $\qquad \square$

Lemma 2.10 *Let $n \geq m \geq 2$ and set M to be the smallest k such that $\binom{k+1}{2} \geq m + n$. The maximum likelihood threshold of the complete bipartite graph $K_{m,n}$ is at most $\min\{M, m+1\}$.*

Proof Since the generic completion rank of $K_{m,n}$ is $m + 1$ in the case $n > \binom{m}{2}$ by Theorem 2.1, the maximum likelihood threshold of $K_{m,n}$ is at most $m + 1$ in this case. So from now on, we assume $n \leq \binom{m}{2}$. In this case, the integer M is at most m. We estimate the dimension of the image of the map

$$
\phi : \begin{cases} \mathbb{R}^n \times O \times \mathbb{R}^m \times \mathrm{Gr}(k, m) & \to & \mathrm{Gr}(k, m+n) \\ (D_n, B, D_m, W) & \mapsto & \left(W \mid W D_m B^t D_n^{-1}\right) \end{cases},
$$

where O is the set of $n \times m$ matrices such that $B^t B = I_m$ and $M \leq k \leq m$. Again, the computation using the Schur complements at the beginning of the proof of Lemma 2.8 shows that the linear spaces in the image of this map are the kernels of matrices of the form

$$
M = \begin{pmatrix} D_m^2 & -D_m B^t D_n \\ -D_n B D_m & D_n^2 \end{pmatrix}.
$$

The dimension of the domain is $n + \sum_{i=1}^{m}(n-i) + m + k(m-k)$ and the dimension of the image is at most $n + \sum_{i=1}^{k}(n-i) + m + k(m-k) - 1$ because B is only determined on a k-dimensional subspace of $\mathbb{R}^m$ and there is a 1-dimensional stabilizer. Indeed,

given $\alpha \in \mathbb{R}^*$, we have $\phi(\mathrm{diag}_n(\alpha)D_n, B, D_m\,\mathrm{diag}_m(\alpha), W) = \phi(D_n, B, D_m, W)$. We compare it to the dimension of $\mathrm{Gr}(k, m+n)$:

$$k(m+n-k) - \left(n + \sum_{i=1}^{k}(n-i) + m + k(m-k) - 1\right) = -m - n + \binom{k+1}{2} + 1.$$

So, if $k \leq m$ and $\binom{k+1}{2} > m+n-1$, then the image of ϕ cannot have non-empty interior in $\mathrm{Gr}(k, m+n)$. In other words, the maximum likelihood threshold of $K_{m,n}$ is at most the smallest k such that $\binom{k+1}{2} \geq m+n$ by Proposition 1.6 (2). □

Proof of Theorem 2.7 The lower bound in Lemma 2.8 matches the upper bound in Lemma 2.10 on the maximum likelihood threshold of $K_{m,n}$. So we conclude that the maximum likelihood threshold of $K_{m,n}$ is the smallest k such that $\binom{k+1}{2} \geq m+n$ if $n \leq \binom{m}{2}$, and $m+1$ if $n > \binom{m}{2}$. □

Corollary 2.11 (a) *The maximum likelihood threshold of $K_{5,5}$ is 4, whereas the generic completion rank is 5.*
   (b) *Asymptotically, the maximum likelihood threshold of $K_{m,m}$ is $2\sqrt{m}$.*
   (c) *The maximum likelihood threshold of $K_{m,n}$ is equal to the generic completion rank of $K_{m,n}$ for $m = 2, 3, 4$ and whenever $n > \binom{m}{2}$.*

The case $m = 2$ in (c) was known before, see [29, Prop. 4.2]

Proof (a) Follows directly from Theorem 2.7. (b) also follows from Theorem 2.7 because $\binom{k+1}{2} \geq 2m$ is quadratic in k and the constant coefficient, after rescaling to make the inequality monic in k, is $4m$. To prove (c), we discuss the several cases. For $m = 2$, we are always in the case $n > \binom{m}{2}$ because $n \geq m$, so the generic completion rank of $K_{2,n}$ is 3 and the maximum likelihood threshold is 3. For $m = 3$, we have two cases. If $n = 3 = \binom{3}{2}$, the generic completion rank is 3 as well as the maximum likelihood threshold. If $n > 3$, the generic completion rank and maximum likelihood threshold are 4. For $m = 4$, the generic completion rank and the maximum likelihood threshold are 4 for $n = 4, 5, 6$ and both are 5 for $n > 6$. □

Remark 2.12 It is essential for the dimension count in the proof of Theorem 2.7 that we are considering positive semidefinite matrices $M \in L_{K_{m,n}}$. For any linear series $W \subset R_1$ of dimension $m-1$, we can find a matrix M of rank 2 such that W is contained in the kernel of M. Namely, choose a linear form $\langle v, -\rangle$ on $\mathbb{R}^m$ that contains the projection of W onto the first m components and a linear form $\langle w, -\rangle$ on $\mathbb{R}^n$ that contains the projection of W onto the last n components. Then the matrix

$$M = \begin{pmatrix} 0 & v^t w \\ w^t v & 0 \end{pmatrix}$$

has rank 2 and contains W in its kernel. Since the diagonal is 0, it cannot be positive semidefinite.

Remark 2.13 The maximum likelihood threshold of the complete bipartite graph $K_{m,m}$ is roughly $2\sqrt{m}$ and its generic completion rank is m. By dropping edges, we can decrease the generic completion rank. Using Theorem 1.4, we know that a subgraph of $K_{m,m}$ of rank roughly $2\sqrt{m}$ can have at most about $4m\sqrt{m}$ edges. So in order for the maximum likelihood threshold and the generic completion rank of G to have a chance to be equal, G needs to be relatively sparse.

3 Non-symmetric Completion

We now relate the general low rank matrix completion problem over $\mathbb{C}$ to the matrix completion problem in the symmetric setup for the special case of bipartite graphs. The low rank matrix completion problem has been extensively studied, especially algorithmically under the assumption that a low rank completion exists (e.g. Netflix Problem). The generic completion rank is a coarse estimate of how reasonable this assumption is for a random partial matrix (i.e., a random matrix with a fixed pattern of known entries).

Let A be a partially specified $m \times n$ matrix. To A, we associate a bipartite graph G on the disjoint union $[m] \cup [n]$, where the two parts index the rows and columns, respectively. We draw the edge $\{i, j\}$ in G if and only if the (i, j)th entry of A is specified. The *non-symmetric generic completion rank* of G is the smallest r such that a generic partially specified $m \times n$ matrix A with the pattern of known entries given by G has a completion of rank r. We want to relate the non-symmetric generic completion rank to the generic completion rank of G, i.e., the generic completion rank in the symmetric setup considered before.

Proposition 3.1 *Let G be a bipartite graph.*

(a) *The non-symmetric generic completion rank of G is at most the symmetric generic completion rank of G.*
(b) *The symmetric generic completion rank of a bipartite graph G is at most the non-symmetric generic completion rank of G plus one.*

Proof Given a G-partial $m \times n$ matrix A, we consider the G-partial symmetric block matrix

$$M' = \begin{pmatrix} D_m & A \\ A^t & D_n \end{pmatrix},$$

where D_m and D_n are matrices whose diagonal entries are specified. To prove (a), let M be a completion of rank r of the G-partial symmetric matrix M'. Then there is a $(m + n) \times r$ matrix U such that $M = UU^t$. Since A is a block of M, such a factorization of M gives a factorization $A = U_1 U_2^t$ of A, where U_1 and U_2 are the appropriate $m \times r$, resp. $n \times r$, blocks of U, which shows (a).

To prove (b), define the projection map $\pi_G^n \colon M_{m \times n} \to \mathbb{A}^{\#E}$ from the set of $m \times n$ matrices to the space indexed by the edges of G mapping a matrix (a_{ij}) to $(a_{ij} : \{i, j\} \in E)$. Let r be the non-symmetric generic completion rank of G so that π_G^n restricted to the variety of matrices of rank at most r is dominant. This implies the

inequality $r(m + n) - r^2 - \#E \geq 0$ by dimension count. It also implies that the symmetric projection map $\pi_G : \mathcal{S}^{m+n} \to \mathbb{A}^{\#E}$ that takes a symmetric matrix (a_{ij}) to $(a_{ij} : \{i, j\} \in E)$ restricted to the variety of symmetric matrices of rank at most r is dominant. Indeed, if a generic G-partial $m \times n$ matrix A has a completion UW^t of rank r, then the symmetric G-partial matrix

$$M = \begin{pmatrix} * & A \\ A^t & * \end{pmatrix}$$

has the rank r completion

$$\begin{pmatrix} U \\ W \end{pmatrix} \begin{pmatrix} U^t & W^t \end{pmatrix}.$$

In particular, the generic fiber dimension of the morphism $\pi_G|_{V_{r+1}^{m+n}} : V_{r+1}^{m+n} \to \mathbb{A}^{\#E}$, where V_{r+1}^{m+n} is the variety of symmetric matrices of rank at most $r + 1$, is $(r + 1)(m + n) - \binom{r+1}{2} - \#E$, because it is also dominant. Using the above inequality $r(m + n) - r^2 - \#E \geq 0$ together with $\binom{r+1}{2} \leq r^2$ for all $r \geq 1$, we get

$$m + n + r(m + n) - \binom{r + 1}{2} - \#E \geq m + n + (r(m + n) - r^2 - \#E) \geq m + n.$$

Fixing the diagonal entries of a partial symmetric matrix in $\mathbb{A}^{\#E}$ imposes linear conditions on the fiber. Since a G-partial symmetric matrix consists of a partial matrix in $\mathbb{A}^{\#E}$ and $m + n$ diagonal entries, this dimension count suggests that we can complete a generic G-partial symmetric matrix to a symmetric matrix of rank $r + 1$. Indeed, this is generically true, because the varieties in question are affine cones of quasi-projective varieties and so the claim follows by projective dimension theory. □

References

1. Agler, J., Helton, J.W., McCullough, S., Rodman, L.: Positive semidefinite matrices with a given sparsity pattern. Linear Algebra Appl. **107**, 101–149 (1988)
2. Barvinok, A.: A Course in Convexity. Graduate Studies in Mathematics, vol. 54. American Mathematical Society, Providence (2002)
3. Ben-David, E.: Sharp lower and upper bounds for the Gaussian rank of a graph. J. Multivariate Anal. **139**, 207–218 (2015)
4. Bernstein, D.I.: Completion of tree metrics and rank-2 matrices (2016). arXiv:1612.06797
5. Blekherman, G.: Positive Gorenstein ideals. Proc. Am. Math. Soc. **143**(1), 69–86 (2015)
6. Blekherman, G., Sinn, R., Velasco, M.: Do sums of squares dream of free resolutions? SIAM J. Appl. Algebra Geom. **1**(1), 175–199 (2017)
7. Bolker, E.D., Roth, B.: When is a bipartite graph a rigid framework? Pacific J. Math. **90**(1), 27–44 (1980)
8. Buhl, S.L.: On the existence of maximum likelihood estimators for graphical Gaussian models. Scand. J. Stat. **20**(3), 263–270 (1993)

9. Choi, M.D., Lam, T.Y., Reznick, B.: Sums of squares of real polynomials. In: Jacob, W., Rosenberg, A. (eds.) *K*-Theory and Algebraic Geometry: Connections with Quadratic Forms and Division Algebras (Santa Barbara, CA, 1992). Proceedings of Symposia in Pure Mathematics, vol. 58, pp. 103–126. American Mathematical Society, Providence (1995)

10. Chua, L., Plaumann, D., Sinn, R., Vinzant, C.: Gram spectrahedra (2016). arXiv:1608.00234

11. Dempster, A.P.: Covariance selection. Biometrics **28**(1), 157–175 (1972)

12. Diestel, R.: Graph Theory. Graduate Texts in Mathematics, vol. 173, 2nd edn. Springer, New York (2000)

13. Dobra, A., Hans, C., Jones, B., Nevins, J.R., Yao, G., West, M.: Sparse graphical models for exploring gene expression data. J. Multivariate Anal. **90**(1), 196–212 (2004)

14. Eisenbud, D., Green, M., Hulek, K., Popescu, S.: Small schemes and varieties of minimal degree. Am. J. Math. **128**(6), 1363–1389 (2006)

15. Flenner, H., O'Carroll, L., Vogel, W.: Joins and Intersections. Springer Monographs in Mathematics. Springer, Berlin (1999)

16. Grone, R., Johnson, C.R., de Sá, E.M., Wolkowicz, H.: Positive definite completions of partial Hermitian matrices. Linear Algebra Appl. **58**, 109–124 (1984)

17. Grone, R., Pierce, S.: Extremal bipartite matrices. Linear Algebra Appl. **131**, 39–50 (1990)

18. Gross, E., Sullivant, S.: The maximum likelihood threshold of a graph (2014). arXiv:1404.6989 (to appear in Bernoulli)

19. Harris, J.: Algebraic Geometry. Graduate Texts in Mathematics. Springer, New York (1995)

20. Hartshorne, R.: Algebraic Geometry. Graduate Texts in Mathematics, vol. 52. Springer, New York (1977)

21. Johnson, C.R., McKee, T.A.: Structural conditions for cycle completable graphs. Discrete Math. **159**(1–3), 155–160 (1996)

22. Krumsiek, J., Suhre, K., Illig, T., Adamski, J., Theis, F.J.: Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. BMC Syst. Biol. **5**(1), 21 (2011)

23. Laurent, M.: On the sparsity order of the graph and its deficiency in chordality. Combinatorica **21**(4), 543–570 (2001)

24. Laurent, M., Varvitsiotis, A.: The Gram dimension of a graph. In: Mahjoub, A.R., et al. (eds.) Combinatorial Optimization. Lecture Notes in Computer Science, vol. 7422, pp. 356–367. Springer, Heidelberg (2012)

25. Lauritzen, S.L.: Graphical Models. Oxford Statistical Science Series, vol. 17. Oxford University Press, New York (1996)

26. Nevo, E.: On embeddability and stresses of graphs. Combinatorica **27**(4), 465–472 (2007)

27. Raymond, J.-L.: Generic rigidity of complete bipartite graphs in R^d. Struct. Topol. **10**, 57–62 (1984)

28. Schäfer, J., Strimmer, K.: An empirical Bayes approach to inferring large-scale gene association networks. Bioinformatics **21**(6), 754–764 (2005)

29. Uhler, C.: Geometry of maximum likelihood estimation in Gaussian graphical models. Ann. Stat. **40**(1), 238–261 (2012)

30. Whittaker, J.: Graphical Models in Applied Multivariate Statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Wiley, Chichester (1990)

31. Zhang, F. (ed.): The Schur Complement and Its Applications. Numerical Methods and Algorithms, vol. 4. Springer, New York (2005)